

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Hiroshi KANAZAWA, et al.

GAU:

SERIAL NO: NEW APPLICATION

EXAMINER:

FILED: HEREWITH

FOR: METHOD AND APPARATUS FOR SUPPRESSING NOISE COMPONENTS CONTAINED IN SPEECH  
SIGNAL

REQUEST FOR PRIORITY

ASSISTANT COMMISSIONER FOR PATENTS  
WASHINGTON, D.C. 20231

SIR:

- ☐ Full benefit of the filing date of U.S. Application Serial Number , filed , is claimed pursuant to the provisions of 35 U.S.C. §120.
- ☐ Full benefit of the filing date of U.S. Provisional Application Serial Number , filed , is claimed pursuant to the provisions of 35 U.S.C. §119(e).
- ☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

COUNTRY

APPLICATION NUMBER

MONTH/DAY/YEAR

JAPAN

2001-017072

January 25, 2001

Certified copies of the corresponding Convention Application(s)

- ☒ are submitted herewith
- ☐ will be submitted prior to payment of the Final Fee
- ☐ were filed in prior application Serial No. filed
- ☐ were submitted to the International Bureau in PCT Application Number  
Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.
- ☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and
- ☐ (B) Application Serial No.(s)
- ☐ are submitted herewith
- ☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,  
MAIER & NEUSTADT, P.C.



Marvin J. Spivak

Registration No. 24,913

C. Irvin McClelland  
Registration Number 21,124



22850



日 本 国 特 許 庁  
JAPAN PATENT OFFICE

J1046 U.S. PTO  
10/054938  
01/25/02

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 1月25日

CERTIFIED COPY OF  
PRIORITY DOCUMENT

出 願 番 号

Application Number:

特願2001-017072

出 願 人

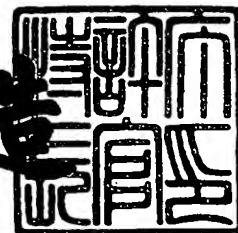
Applicant(s):

株式会社東芝

2001年12月21日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

及 川 耕 造



【書類名】 特許願

【整理番号】 A000006986

【提出日】 平成13年 1月25日

【あて先】 特許庁長官 殿

【国際特許分類】 G10L 3/02

【発明の名称】 音声信号の雑音抑圧方法及び音声認識装置

【請求項の数】 9

【発明者】

    【住所又は居所】 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研  
    究開発センター内

    【氏名】 押切 正浩

【発明者】

    【住所又は居所】 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研  
    究開発センター内

    【氏名】 金澤 博史

【特許出願人】

    【識別番号】 000003078

    【氏名又は名称】 株式会社 東芝

【代理人】

    【識別番号】 100058479

    【弁理士】

    【氏名又は名称】 鈴江 武彦

    【電話番号】 03-3502-3181

【選任した代理人】

    【識別番号】 100084618

    【弁理士】

    【氏名又は名称】 村松 貞男

【選任した代理人】

    【識別番号】 100068814

【弁理士】

【氏名又は名称】 坪井 淳

【選任した代理人】

【識別番号】 100092196

【弁理士】

【氏名又は名称】 橋本 良郎

【選任した代理人】

【識別番号】 100091351

【弁理士】

【氏名又は名称】 河野 哲

【選任した代理人】

【識別番号】 100088683

【弁理士】

【氏名又は名称】 中村 誠

【選任した代理人】

【識別番号】 100070437

【弁理士】

【氏名又は名称】 河井 将次

【手数料の表示】

【予納台帳番号】 011567

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声信号の雑音抑圧方法及び音声認識装置

【特許請求の範囲】

【請求項1】 入力音声信号に含まれる雑音成分を抑圧する音声信号の雑音抑圧方法において、

前記入力音声信号を所定フレーム長で周波数分析して入力スペクトルを算出するステップと、

前記雑音成分のスペクトルを推定して推定雑音スペクトルを得るステップと、

前記推定雑音スペクトルに所定のスペクトル減算係数を乗じるステップと、

前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクシヨンスペクトルを得るステップと、

前記サブトラクシヨンスペクトルのクリッピングを行って音声スペクトルを求めるステップと、

前記音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑化して修正することにより、前記雑音成分が抑圧された出力音声信号を得るステップと

を有する音声信号の雑音抑圧方法。

【請求項2】 入力音声信号に含まれる雑音成分を抑圧する音声信号の雑音抑圧方法において、

前記入力音声信号を所定フレーム長で周波数分析することによりスペクトルを算出して入力スペクトルを得るステップと、

前記雑音成分のスペクトルを推定して推定雑音スペクトルを得るステップと、

前記推定雑音スペクトルのスペクトル傾きを求めるステップと、

前記推定雑音スペクトルに前記スペクトル傾きの度合いによって決定されるスペクトル減算係数を乗じるステップと、

前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクシヨンスペクトルを得るステップと、

前記サブトラクシヨンスペクトルのクリッピングを行って音声スペクトルを求めることにより、前記雑音成分が抑圧された出力音声信号を得るステップと

を有する音声信号の雑音抑圧方法。

【請求項3】 入力音声信号に含まれる雑音成分を抑圧する音声信号の雑音抑圧方法において、

前記入力音声信号を所定フレーム長で周波数分析することによりスペクトルを算出して入力スペクトルを得るステップと、

前記雑音成分のスペクトルを推定して推定雑音スペクトルを得るステップと、

前記推定雑音スペクトルのスペクトル傾きを求めるステップと、

前記推定雑音スペクトルに前記スペクトル傾きの度合いによって決定されるスペクトル減算係数を乗じるステップと、

前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクションスペクトルを得るステップと、

前記サブトラクションスペクトルのクリッピングを行って音声スペクトルを求めるステップと、

前記音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑化して修正することにより、前記雑音成分が抑圧された出力音声信号を得るステップと

を有する音声信号の雑音抑圧方法。

【請求項4】 入力音声信号に含まれる雑音成分を抑圧する音声信号の雑音抑圧装置において、

前記入力音声信号を所定フレーム長で周波数分析して入力スペクトルを得る周波数分析手段と、

前記雑音成分のスペクトルを推定して推定雑音スペクトルを得る雑音スペクトル推定手段と、

前記推定雑音スペクトルにスペクトル減算係数を乗じる乗算手段と、

前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクションスペクトルを得る減算手段と、

前記サブトラクションスペクトルのクリッピングを行って音声スペクトルを求めるクリッピング手段と、

前記音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑

化して修正することにより、前記雑音成分が抑圧された出力音声信号を得る修正手段と

を有する音声信号の雑音抑圧装置。

【請求項5】 入力音声信号に含まれる雑音成分を抑圧する音声信号の雑音抑圧装置において、

前記入力音声信号を所定フレーム長で周波数分析して入力スペクトルを得る周波数分析手段と、

前記雑音成分のスペクトルを推定して推定雑音スペクトルを得る雑音スペクトル推定手段と、

前記推定雑音スペクトルのスペクトル傾きを求めるスペクトル傾き算出手段と

前記スペクトル傾きの度合いによって決定されるスペクトル減算係数を算出するスペクトル減算係数算出手段と、

前記推定雑音スペクトルにスペクトル減算係数を乗じる乗算手段と、

前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクションスペクトルを得る減算手段と、

前記サブトラクションスペクトルのクリッピングを行って音声スペクトルを求めることにより、前記雑音成分が抑圧された出力音声信号を得るクリッピング手段と

を有する音声信号の雑音抑圧装置。

【請求項6】 入力音声信号に含まれる雑音成分を抑圧する音声信号の雑音抑圧装置において、

前記入力音声信号を所定フレーム長で周波数分析して入力スペクトルを得る周波数分析手段と、

前記雑音成分のスペクトルを推定して推定雑音スペクトルを得る雑音スペクトル推定手段と、

前記推定雑音スペクトルのスペクトル傾きを求めるスペクトル傾き算出手段と

前記スペクトル傾きの度合いによって決定されるスペクトル減算係数を算出す

るスペクトル減算係数算出手段と、

前記推定雑音スペクトルにスペクトル減算係数を乗じる乗算手段と、

前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクシヨンスペクトルを得る減算手段と、

前記サブトラクシヨンスペクトルのクリッピングを行って音声スペクトルを求めるクリッピング手段と、

前記音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑化して修正することにより、前記雑音成分が抑圧された出力音声信号を得る音声スペクトル修正手段と

を有する音声信号の雑音抑圧装置。

【請求項7】 前記スペクトル修正手段は、前記音声スペクトルの各々を周波数軸上及び時間軸上の少なくとも一方で近傍に存在する音声スペクトルを用いて平滑化する請求項4または6記載の音声信号の雑音抑圧装置。

【請求項8】 前記スペクトル修正手段は、周波数軸上及び時間軸上の少なくとも一方で前記音声スペクトルと所定の関数とで畳み込みを行う請求項4または6記載の音声信号の雑音抑圧装置。

【請求項9】 請求項4乃至8のいずれか1項記載の雑音抑圧装置と、前記雑音抑圧装置からの出力音声信号について認識処理を行う音声認識部とを有する音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声の雑音成分を抑圧する雑音抑圧方法及び装置並びに雑音抑圧装置を備えた音声認識装置に関する。

【0002】

【従来の技術】

雑音環境下での音声を聞きやすくしたり、音声認識の認識率を向上させるために、音声信号に含まれる背景雑音などの雑音成分を抑圧する技術が用いられる。従来の雑音抑圧技術の中で、比較的少ない計算量で効果が得られる方法として、



例えば文献1：“S.Fboll, Suppression of acoustic noise in speech using spectral subtraction, IEEE Transactions on Acoustics, Speech and Signal processing, Vol.Assp-27, No.2, April 1979, pp.113-120”に記載されたスペクトルサブトラクション法が知られている。

#### 【0003】

スペクトルサブトラクション法は、入力音声信号を周波数分析してパワーもしくは振幅のスペクトル（入力スペクトルという）を求め、雑音区間で推定された推定雑音スペクトルに所定の係数（スペクトル減算係数） $\alpha$ を乗じ、このスペクトル減算係数 $\alpha$ を乗じた後の推定雑音スペクトルを入力スペクトルから減じることによって雑音成分を抑圧する方法である。実際には、入力スペクトルから推定雑音スペクトルを減じた後のスペクトルが零もしくは零に近い所定値より小さくなったとき、その所定値をクリッピングレベルとしてクリッピングを行うことにより、最終的に雑音成分が抑圧された出力音声信号を得る。

#### 【0004】

図1及び図2を用いて、スペクトルサブトラクション法により雑音を抑圧する様子を説明する。図1は、入力音声信号の音声区間を所定フレーム長で周波数分析して得られる入力スペクトル（実線）と、推定雑音スペクトル（点線）、及び入力スペクトルから推定雑音スペクトルを減算し、さらにクリッピングを行った後の出力スペクトル（破線）を表す。図2は、雑音が重畳していないクリーン条件で入力音声信号の同一区間をスペクトル分析した結果を示す。

#### 【0005】

入力スペクトルを $X(m)$ 、推定雑音スペクトルを $N(m)$ としたとき、出力スペクトル $Y(m)$ は、

$$Y(m) = \max(X(m) - \alpha N(m), T_{cl} \cdot X(m))$$

と表される。ここで、 $\max()$ は最大値を返す関数、 $T_{cl}$ はクリッピング係数を表し、 $m$ は周波数に対応するインデックスを表す。

また、スペクトル減算係数 $\alpha$ を1よりも大きな値に設定し、推定雑音スペクトルの元々の値よりも大きな量を入力スペクトルから減算する方法がとられることがある。これは一般に過減算(Over-subtraction)と呼ばれ、音声認識で有効な手

法である。

【0006】

【発明が解決しようとする課題】

上述したスペクトルサブトラクション法により雑音を抑圧する場合、理想的には出力スペクトル $Y(m)$ は図2のクリーン条件下のスペクトルに近づくことが望まれる。しかし、実際には図1に示すように出力スペクトル $Y(m)$ は、ホルマント部にピーク状のスペクトルがいくつか残り、それ以外のスペクトルは大きく減衰してしまうため、ホルマントの形状を正確に表すことができなくなる（第1の問題点）。

【0007】

この第1の問題点は、次のことに起因する。入力スペクトル $X(m)$ と推定雑音スペクトル $N(m)$ との関係が $X(m) - \alpha N(m) > T_{cl} \cdot X(m)$ を満たす場合には、出力スペクトル $Y(m)$ に $X(m) - \alpha N(m)$ の値が与えられる（図1の矢印A）。この条件を満たさない場合には、クリッピング係数が乗じられたスペクトル $T_{cl} \cdot X(m)$ が出力スペクトル $Y(m)$ として出力される。スペクトルサブトラクションの効果を出すには、クリッピング係数 $T_{cl}$ を0.01のような非常に小さい値に設定する必要がある、このために第1の問題点が生じる。

【0008】

一方、推定雑音スペクトルの形状によっては、本来スペクトルピークが残るべき部分でスペクトルピークが消失してしまうことがある（第2の問題点）。図3は、図1と同一区間の入力音声信号に中域のパワーが相対的に大きい雑音成分が重畳したときの入力スペクトルを表している。入力スペクトルと雑音スペクトルがこのような関係にあるとき、矢印Bに本来あるべきスペクトルピークが消失してしまい、図3の場合は図2の第2ホルマント $F_2$ を表す情報が消失している。この結果、音声認識における認識率が低下する。

【0009】

効果的なスペクトルサブトラクション法を実現するためには、雑音スペクトルの正確な推定が不可欠である。一般に、雑音スペクトルの推定においては入力音声信号の非音声区間を周波数分析し、その平均値を推定雑音スペクトルとする。

しかし、雑音環境下で非音声区間を正確に判定することは非常に困難であり、音声区間のスペクトルを用いて推定雑音スペクトルを算出してしまうことが頻繁に生じてしまう。

#### 【0010】

音声区間の最初（語頭）では、子音など比較的スペクトル特性が高域に偏る音韻が出現する場合が多く、このため推定雑音スペクトルの値が高域ほど実際の雑音スペクトルより大きくなる。そのため、必要以上に推定雑音スペクトルが入力スペクトルから減算されてしまい、正しい雑音抑圧ができなくなる（第3の問題点）。

#### 【0011】

図4及び図5は、本来の雑音スペクトルと、非音声区間の判定に失敗し子音のスペクトルを用いて雑音スペクトルを推定してしまった場合を表す図である。図4は、本来の雑音スペクトルの高域の振幅が大きい場合、図5は本来の雑音スペクトルの高域の振幅が小さい場合を表す。両者を比較すると、雑音スペクトルの形状によって影響の受け方は異なり、雑音スペクトルの高域振幅が小さいほど影響を受けやすいことが分かる。すなわち、推定雑音スペクトルの高域振幅が小さいほど雑音スペクトルの推定誤差が生じ、必要以上に入力スペクトルから推定雑音スペクトルが減算される傾向が強くなる。

#### 【0012】

上述した3つの問題点は、主として推定雑音スペクトルの信頼性が低い場合、雑音スペクトルの特性が変動した場合、音声信号の複素スペクトルと雑音成分の複素スペクトルの位相が大きく異なる場合などに生じ、音声認識における認識率低下の原因となる。

#### 【0013】

##### 【発明が解決しようとする課題】

上述したように、従来の雑音抑圧技術では（1）出力の音声スペクトルが入力音声信号のホルマント形状を正確に表すことができなくなる、（2）推定雑音スペクトルの形状によっては、本来スペクトルピークが残るべき部分でスペクトルピークが消失してしまう、及び（3）雑音スペクトルの推定誤差により必要以上

に推定雑音スペクトルが入力スペクトルから減算される、といった問題点があるため、的確な雑音抑圧を行うことができず、音声認識の前処理に用いた場合、認識率の向上にあまり有効でなかった。

## 【0014】

本発明の主たる目的は、入力音声信号に含まれる雑音成分を音声信号のスペクトルを損なうことなく抑圧できる雑音抑圧方法及び装置を提供することにある。

## 【0015】

さらに、本発明の他の目的は、雑音抑圧処理を音声認識の前処理として適用することにより、雑音環境下でも高い認識率が得られる音声認識装置を提供することにある。

## 【0016】

## 【課題を解決するための手段】

上記の課題を解決するため、本発明の一つの態様においては、入力音声信号を所定フレーム長で周波数分析して得られる入力スペクトルから、入力音声信号に含まれる雑音成分のスペクトルを推定して得られる推定雑音スペクトルに所定のスペクトル減算係数を乗じた後の推定雑音スペクトルを減じてサブトラクションスペクトルを生成し、このサブトラクションスペクトルのクリッピングを行って雑音成分が抑圧された音声スペクトルを求めた後、この音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑化して修正することにより、雑音成分が抑圧された出力音声信号を生成する。

## 【0017】

音声スペクトルの平滑化は、例えば音声スペクトルの各々を周波数軸上及び時間軸上の少なくとも一方で近傍に存在する音声スペクトルを用いて行われるか、あるいは周波数軸上及び時間軸上の少なくとも一方で音声スペクトルと所定の関数とで畳み込みを行うことにより行われる。

## 【0018】

サブトラクションスペクトルのクリッピングを行って得られた音声スペクトルを周波数軸上で平滑化して修正することにより、音声スペクトルはクリーン条件でのスペクトルの概形に近づくため、雑音成分が除去され、かつ入力音声信号の

ホルマント形状を正確に反映した出力音声信号が得られる。

【0019】

また、サブトラクシヨンスpekトルのクリッピングを行って得られた音声pekトルを時間軸上で平滑化して修正することによって、雑音成分が除去され、かつクリッピングにより消失したpekトルピークが復元された出力音声信号が得られる。

【0020】

本発明の他の態様によると、推定雑音pekトルのpekトル傾きを求め、推定雑音pekトルにpekトル傾きの度合いによって決定されるpekトル減算係数を乗じ、入力pekトルから該pekトル傾きの度合いによって決定されたpekトル減算係数を乗算した後のpekトルを減じてサブトラクシヨンスpekトルを生成し、このサブトラクシヨンスpekトルのクリッピングを行って音声pekトルを求めることにより、雑音成分が抑圧された出力音声信号を生成する。

【0021】

このようにすると、推定雑音pekトルの信頼性が低い場合でも、推定雑音pekトルの傾きの程度によってpekトル減算係数が定まるため、効果的な雑音抑圧が実現できる。

【0022】

本発明のさらに別の態様では、上述した二つの態様を組み合わせ、入力pekトルから該pekトル傾きの度合いによって決定されたpekトル減算係数を乗算した後のpekトルを減じてサブトラクシヨンスpekトルを生成し、このサブトラクシヨンスpekトルのクリッピングを行って音声pekトルを求めた後、この音声pekトルの各々を周波数軸上及び時間軸上の少なくとも一方で平滑化して修正することにより、雑音成分が抑圧された出力音声信号を生成する。

【0023】

さらに、本発明によると上述した雑音抑圧装置を音声認識部の前段に配置することにより認識率を向上させた音声認識装置が提供される。

【0024】

【発明の実施の形態】

以下、図面を参照して本発明の実施の形態について説明する。

〔第 1 の実施形態〕

図 6 は、本発明の第 1 の実施形態に係る雑音抑圧装置の構成を示している。また、図 7 は本実施形態における雑音抑圧処理の流れを示している。図 6 及び図 7 に示すように、音声入力端子 1 1 には、音声信号が所定フレーム長に分割されたフレーム単位で入力され、まず周波数分析部 1 2 によって周波数分析が行われる（ステップ S 1 1）。周波数分析部 1 2 では、以下のようにして入力音声信号のスペクトル（入力スペクトル）が算出される。

【 0 0 2 5 】

まず、フレーム単位の音声信号についてハミング窓を用いて窓掛けを行い、その後離散フーリエ変換（DFT）を施す。DFT を施すことにより得られる複素スペクトルをパワースペクトルもしくは振幅スペクトルに変換し、これを入力スペクトル  $X(i, m)$  とする。ここで、 $i$  はフレーム番号を表し、 $m$  は周波数に対応するインデックスを表す。本実施形態では、スペクトルとして振幅スペクトルを用いた例について説明を行うが、パワースペクトルを用いても構わない。以後、特に断りのない限りスペクトルとは振幅スペクトルを意味することにする。

【 0 0 2 6 】

次に、雑音スペクトル推定部 1 3 に保存されている推定雑音スペクトル  $N(i, m)$  に、スペクトル減算係数格納部 1 4 に格納されているスペクトル減算係数  $\alpha$  が乗算器 1 5 により乗じられる（ステップ S 1 2）。

【 0 0 2 7 】

次に、減算器 1 6 により乗算器 1 5 から出力されるスペクトルが次式のように入力スペクトル  $X(i, m)$  から差し引かれ（ステップ S 1 3）、 $Y(i, m)$  なるスペクトル（サブトラクションスペクトル）が生成される。

【数 1】

$$Y(i, m) = X(i, m) - \alpha \cdot N(i, m) \quad (1)$$

【 0 0 2 8 】

減算器 1 6 からのサブトラクションスペクトル  $Y(i, m)$  はクリッピング部 1 7 に入力され、次式に示すようにサブトラクションスペクトル  $Y(i, m)$  が閾値

$\gamma \cdot X(i, m)$ よりも小さい場合には、 $\gamma \cdot X(i, m)$ で置き換えられることによってクリッピングが行われ、音声スペクトルが求められる（ステップS14）。このクリッピングは、音声スペクトルが負値になることを回避するために行われる。

【数2】

$$Y(i, m) = \gamma \cdot X(i, m) \text{ if } X(i, m) - \alpha \cdot N(i, m) < \gamma \cdot X(i, m) \quad (2)$$

ここで、 $\gamma$ は零もしくは零に近い小さい定数を表し、本実施形態では $\gamma = 0.01$ とする。

【0029】

次に、スペクトル修正部18によってクリッピング後のスペクトルである音声スペクトル $Y(i, m)$ の修正が行われる（ステップS15）。フレーム番号 $i$ 、周波数 $m$ の音声スペクトル $Y(i, m)$ を修正して得られる修正後のスペクトル（修正スペクトル）を $Y'(i, m)$ と表すことにする。この修正スペクトル $Y'(i, m)$ が出力音声信号として音声出力端子19から出力される。

【0030】

スペクトル修正部18における音声スペクトル $Y(i, m)$ の修正法としては、以下に示すように音声スペクトル $Y(i, m)$ を周波数軸上で近傍の音声スペクトルを用いて修正する方法と、時間軸上で近傍の音声スペクトルを用いて修正する方法とがある。なお、ここでは明示的に説明は行わないが、音声スペクトル $Y(i, m)$ を周波数軸上で近傍の音声スペクトル及び時間軸上で近傍の音声スペクトルの両方を用いて修正するようにしてもよい。

【0031】

（音声スペクトルを周波数軸上で近傍のスペクトルを用いて修正する方法）

最初に、音声スペクトルを周波数軸上で近傍の音声スペクトルを用いて修正する方法について説明する。修正スペクトル $Y'(i, m)$ は、音声スペクトル $Y(i, m)$ の周波数軸上で近傍の音声スペクトル $Y(i, m+k)$  ( $k = -K1, -K1+1, \dots, K2$ )を用いて算出される。ここで $K1, K2$ は正の定数を表す。具体的には、修正スペクトル $Y'(i, m)$ は、

【数3】

$$Y'(i, m) = \max(Y(i, m+k)) \quad (k = -K1, -K1+1, \dots, K2) \quad (3)$$

として求められる。ここで  $\max()$  は最大値を出力する関数を表す。この方法では、音声スペクトル  $Y(i, m)$  を周波数軸上で近傍のスペクトル  $Y(i, m+k)$  の中の最大値で置き換えて修正スペクトル  $Y'(i, m)$  とすることを意味する。この方法による効果を図 8 を用いて説明する。図 8 では  $K1=K2=1$  としている。

## 【0032】

図 8 において、実線は修正前の音声スペクトル  $Y(i, m)$ 、点線は上述の方法で修正を行った後に得られる修正スペクトル  $Y'(i, m)$  をそれぞれ表し、また破線は雑音が重畳していないクリーン条件での音声スペクトルを表す。この図を見ると、修正を行うことにより音声スペクトルの平滑化がなされ、よりクリーン条件のスペクトルの概形に近づくことが分かる。従って、前述した第 1 の問題点を解消することができる。

## 【0033】

この効果により、本実施形態による雑音抑圧処理を後述する音声認識部の前処理として適用すれば、認識率の改善を実現できる。一般的に、音声認識ではスペクトルの概形の情報から特徴量を算出することを基本としているため、本実施形態による雑音抑圧処理は非常に有効である。

## 【0034】

この方法の変形として、次式のように 1 以下の正定数  $\beta$  を用いて修正スペクトル  $Y'(i, m)$  を生成してもよく、この場合も同様の効果が得られる。

## 【数 4】

$$Y'(i, m) = \max(\beta^{|k|} \cdot Y(i, m+k)) \quad (k = -K1, -K1+1, \dots, K2) \quad (4)$$

## 【0035】

また、音声スペクトル  $Y(i, m)$  を所定の関数  $h(j)$  と畳み込みを行って修正スペクトル  $Y'(i, m)$  を生成する方法を用いてもよい。この方法は次式で表される。

## 【数 5】

$$Y'(i, m) = \sum_{j=-(J-1)/2}^{(J-1)/2} Y(i, m-j) \cdot h(j+(J-1)/2) \quad (5)$$

ここで、 $J$  は関数  $h(j)$  の要素数を表す。関数  $h(j)$  としては、 $h(j)$  の中心が最大値となる凸関数、例えば  $h(j) = \{0.1, 0.4, 0.7, 1.0, 0.7, 0.4, 0.1\}$  のよ



うな関数が適当である。

【0036】

図9に、この方法による音声スペクトルの修正の様子を示す。図9では図8と同様、実線は修正前の音声スペクトル $Y(i, m)$ 、点線は修正スペクトル $Y'(i, m)$ 、破線は雑音が重畳していないクリーン条件での音声スペクトルを表す。この方法でも上述の方法と同様に音声スペクトルの平滑化がなされ、クリーン条件の音声スペクトルの概形に近づくことが分かる。よって、第1の問題点を解消することができる。

【0037】

(音声スペクトルを時間軸上で近傍のスペクトルを用いて修正する方法)

次に、音声スペクトル $Y(i, m)$ を時間軸上で近傍の音声スペクトルを用いて修正する方法について説明する。修正スペクトル $Y'(i, m)$ は、音声スペクトル $Y(i, m)$ の時間軸上で近傍のスペクトル $Y(i+k, m)$  ( $k = -K1, -K1+1, \dots, K2$ )を用いて算出される。具体的には、修正スペクトル $Y'(i, m)$ は次式で求められる。

【数6】

$$Y'(i, m) = \max(Y(i+k, m)) \quad (k = -K1, -K1+1, \dots, K2) \quad (6)$$

【0038】

この効果を図10を用いて説明する。図10は、音声スペクトル $Y(i, m)$ に本来あるべき第2ホルマントが雑音によって消失した例を表している。 $K1=K2=1$ として修正を行う場合、この例では、 $Y'(i-1, m)$ で第2ホルマントに対応するスペクトルピークが存在しているため、前述したような修正を行うことにより消失したスペクトルピークを復元することができる。これにより、第2の問題点を解消することができる。

【0039】

この方法の変形として、次式のように1以下の正定数 $\beta$ を用いて修正スペクトル $Y'(i, m)$ を生成してもよく、この場合も同様の効果が得られる。

【数7】

$$Y'(i, m) = \max(\beta^{|k|} \cdot Y(i+k, m)) \quad (k = -K1, -K1+1, \dots, K2) \quad (7)$$

## 【0040】

スペクトルピークが消失するか否かは、音声信号と雑音成分の位相関係に依存する。雑音成分の位相はランダムとみなすことができるため、ある時刻ではスペクトルピークが消失しても別の時刻ではスペクトルピークが残る可能性がある。すなわち、長い時間スペクトルを観測するほど、つまり $K1, K2$ を大きくとるほど、スペクトルピークが復元される可能性は高くなる。しかし、余りに長い時間観測すると、異なる音韻で修正される危険性もある。そのため、適切な $K1, K2$ を設定する必要がある。

## 【0041】

また、音声スペクトル $Y(i, m)$ を所定の関数 $h(j)$ と畳み込みを行って修正スペクトル $Y'(i, m)$ を生成する方法を用いてもよい。この方法は次式で表される。

【数8】

$$Y'(i, m) = \sum_{j=-(J-1)/2}^{(J-1)/2} Y(i-j, m) \cdot h(j + (J-1)/2) \quad (8)$$

ここで、 $J$ は関数 $h(j)$ の要素数を表す。関数 $h(j)$ としては、 $h(j)$ の中心が最大値となる凸関数、例えば $h(j) = \{0.1, 0.4, 0.7, 1.0, 0.7, 0.4, 0.1\}$ のような関数が適当である。

## 【0042】

別の方法として、未来のスペクトルを用いずに、現在から過去にかけてのスペクトルのみを用いて修正する、つまり $K2=0$ とする方法もある。この方法では、現在から過去のスペクトルのみを用いているため、時間的な遅延が生じないという利点がある。

## 【0043】

さらに別の方法として、AR (Autoregressive: 自己回帰) 型フィルタによって音声スペクトルの修正を行う方法がある。この場合、修正スペクトル $Y'(i, m)$ は次式で表される。

【数9】

$$Y'(i, m) = Y(i, m) + \sum_{j=1}^J \alpha_{ar}(j) \cdot Y'(i-j, m) \quad (9)$$

ここで $\alpha_{mr}$ はフィルタ係数、 $J$ はフィルタ次数をそれぞれ表す。

【0044】

同様に、MA (Moving Average: 移動平均) 型フィルタによって音声スペクトルの修正を行う方法もあり、この場合の修正スペクトル $Y'(i,m)$ は次式で表される。

【数10】

$$Y'(i,m) = Y(i,m) + \sum_{j=1}^J \alpha_{ma}(j) \cdot Y(i-j,m) \quad (10)$$

ここで $\alpha_{ma}$ はフィルタ係数、 $J$ はフィルタ次数をそれぞれ表す。これらは実現の方法は異なるものの、消失したスペクトルピークを復元して第2の問題点を解消するという意味で、同様の効果が得られる。さらに、以上説明した修正法を組み合わせて用いてもよい。

【0045】

〔第2の実施形態〕

図11に、本発明の第2の実施形態に係る雑音抑圧装置の構成を示す。図11において、図6と同一部分に同一符号を付して説明すると、本実施形態ではスペクトル傾き算出部21が追加されている。スペクトル傾き算出部21では、雑音スペクトル推定部13で求められた推定雑音スペクトルの傾きが算出される。このスペクトル傾きに基づいて、スペクトル減算係数算出部22でスペクトル減算係数 $\alpha$ が算出され、乗算器15に与えられる。本実施形態では、スペクトル減算係数 $\alpha$ として、周波数毎に異なる値を求めるため、以後 $\alpha(m)$ と表記することにする。

【0046】

以下、図12を用いて本実施形態における雑音抑圧処理の流れを説明する。

まず、第1の実施形態と同様に、周波数分析部11によつて入力音声信号の周波数分析を行う(ステップS21)。次に、スペクトル傾き算出部28において推定雑音スペクトル $N(i,m)$ の傾きを算出するために、まず低域部と高域部のスペクトルの比を求める(ステップS22)。このスペクトル比 $r$ は、次式で表される。

【数 1 1】

$$r = \sqrt{\frac{\sum_{m \in FH} N^2(i, m)}{\sum_{m \in FL} N^2(i, m)}} \quad (11)$$

ここで、FLは低域部に属する周波数のインデックスの集合、FHは高域部に属する周波数インデックスの集合を表す。

【0047】

次に、スペクトル減算係数算出部22ではスペクトル比 $r$ を用いてスペクトル減係数 $\alpha(m)$ を求める(ステップS23)。本実施形態では、前述した第3の問題点の観点から、スペクトル比 $r$ が大きいほどスペクトル減算係数 $\alpha(m)$ を小さく、言い換えればスペクトル比 $r$ が小さいほどスペクトル減算係数 $\alpha(m)$ を大きく、また周波数が高いほどスペクトル減算係数 $\alpha(m)$ を小さく、言い換えれば周波数が低いほどスペクトル減算係数 $\alpha(m)$ を大きく設定する。

【0048】

すなわち、スペクトル減算係数 $\alpha(m)$ は、次式のようにスペクトル比 $r$ と周波数インデックス $m$ の関数として表される。

【数 1 2】

$$\alpha(m) = \max(0.0, \min(F(r, m), \alpha_c)) \quad (12)$$

【0049】

関数 $F(r, m)$ は、スペクトル比 $r$ に対して単調減少、周波数インデックス $m$ に対して単調減少となる特徴を持つ。また、関数 $F(r, m)$ の出力は、0.0と $\alpha_c$ の範囲に入るように処理される。ここで、 $\alpha_c$ は最大スペクトル減算係数を表し、例えば $\alpha_c = 2.0$ のように予め設定されている。スペクトル減算係数 $\alpha(m)$ をこのように算出することで、第3の問題点の影響を軽減することができる。

【0050】

関数 $F(r, m)$ の一つの例は、例えば次式で与えられる。この式は上述した条件を満たしている。ここで、 $M$ は最高周波数に対応するインデックスを表す。

【数 13】

$$F(r, m) = \alpha \cdot (1.0 - r \cdot \frac{m}{M-1}) \quad (13)$$

【0051】

次に、乗算器15において雑音スペクトル推定部13で得られた推定雑音スペクトルに、ステップS23で算出されたスペクトル減算係数 $\alpha(m)$ を乗じる（ステップS24）。次に、減算器16において入力スペクトルからスペクトル減算係数 $\alpha(m)$ を乗算した後の推定雑音スペクトルを減算し（ステップS25）、さらに減算後のスペクトルのクリッピングを行うことにより（ステップS26）、雑音成分が抑圧された出力音声信号を得る。

【0052】

【第3の実施形態】

図13に、本発明の第3の実施形態に係る雑音抑圧装置の構成を示す。本実施形態は、第1の実施形態と第2の実施形態を組み合わせた構成、すなわち第2の実施形態である図11のクリッピング部17の後段に、第1の実施形態である図6に示したスペクトル修正部18を配置した構成となっている。このような構成により、本実施形態によると第1及び第2の実施形態の両者の効果を組み合わせた効果が得られる。

【0053】

本実施形態では、図14に処理の流れを示したように、まず入力音声信号を所定フレーム長で周波数分析して入力スペクトルを求め（ステップS31）、推定雑音スペクトルのスペクトル比を求める（ステップS32）。次に、スペクトル減算係数 $\alpha(m)$ を求め（ステップS33）、推定雑音スペクトルにスペクトル減算係数 $\alpha(m)$ を乗じる（ステップS34）。次に、入力スペクトルからスペクトル減算係数 $\alpha(m)$ を乗算した後の推定雑音スペクトルを減算し（ステップS35）、減算後のスペクトルのクリッピングを行う（ステップS36）。最後に、クリッピング後のスペクトルを修正して修正スペクトルを求め（ステップS37）、出力音声信号を得る。

【0054】

#### 【第4の実施形態】

図15は、本発明の第4の実施形態として本発明を音声認識装置に適用した例を示している。図15において、音声入力端子11からの入力音声信号は、雑音抑圧部31に入力され、雑音成分が抑圧される。雑音抑圧部31から音声出力端子19に出力される出力音声信号は、音声認識部32に入力される。音声認識部32は、雑音抑圧部31から出力される音声信号に対して音声認識処理を行い、出力端子20へ認識結果を出力する。

#### 【0055】

ここで、雑音抑圧部31は第1～第3の実施形態で説明したいずれかの雑音抑圧装置である。例えば、雑音抑圧部31が第3の実施形態で説明した雑音抑圧装置であるとする、図13のスペクトル修正部18から修正スペクトル $Y'(i,m)$ が出力され、これが音声信号として音声出力端子19から音声認識部32に入力される。音声認識部32では、修正スペクトル $Y'(i,m)$ を基に音声信号の特徴量を求め、この特徴量に対して所定の辞書に含まれる候補の中で最も類似している候補を認識結果として求め、出力端子20へ出力する。

#### 【0056】

このように本実施形態によれば、本発明に基づく雑音抑圧装置を音声認識の前処理部に用いることにより、高い認識率を実現することができる。

#### 【0057】

上述した本発明による音声信号の雑音抑圧処理は、パーソナルコンピュータやワークステーションのようなコンピュータを用いてソフトウェアにより実行することが可能である。従って、本発明によれば以下に挙げるようなプログラムを格納したコンピュータ読み取り可能な記録媒体、あるいはプログラムを提供することができる。

#### 【0058】

(1) 入力音声信号に含まれる雑音成分を抑圧する処理をコンピュータに実行させるためのプログラムまたは該プログラムを格納したコンピュータ読み取り可能な記録媒体であって、前記入力音声信号を所定フレーム長で周波数分析して入力スペクトルを算出する処理と、前記雑音成分のスペクトルを推定して推定雑音

スペクトルを得る処理と、前記推定雑音スペクトルに所定のスペクトル減算係数を乗じる処理と、前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクションスペクトルを得る処理と、前記サブトラクションスペクトルのクリッピングを行って音声スペクトルを求める処理と、前記音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑化して修正することにより、前記雑音成分が抑圧された出力音声信号を得る処理とをコンピュータに実行させるためのプログラムまたは該プログラムを格納したコンピュータ読み取り可能な記録媒体。

## 【 0 0 5 9 】

(2) 入力音声信号に含まれる雑音成分を抑圧する処理をコンピュータに実行させるためのプログラムまたは該プログラムを格納したコンピュータ読み取り可能な記録媒体であって、前記入力音声信号を所定フレーム長で周波数分析することによりスペクトルを算出して入力スペクトルを得る処理と、前記雑音成分のスペクトルを推定して推定雑音スペクトルを得る処理と、前記推定雑音スペクトルのスペクトル傾きを求める処理と、前記推定雑音スペクトルに前記スペクトル傾きの度合いによって決定されるスペクトル減算係数を乗じる処理と、前記入力スペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクションスペクトルを得る処理と、前記サブトラクションスペクトルのクリッピングを行って音声スペクトルを求めることにより、前記雑音成分が抑圧された出力音声信号を得る処理とをコンピュータに実行させるためのプログラムまたは該プログラムを格納したコンピュータ読み取り可能な記録媒体。

## 【 0 0 6 0 】

(3) 入力音声信号に含まれる雑音成分を抑圧する処理をコンピュータに実行させるためのプログラムまたは該プログラムを格納したコンピュータ読み取り可能な記録媒体であって、前記入力音声信号を所定フレーム長で周波数分析することによりスペクトルを算出して入力スペクトルを得る処理と、前記雑音成分のスペクトルを推定して推定雑音スペクトルを得る処理と、前記推定雑音スペクトルのスペクトル傾きを求める処理と、前記推定雑音スペクトルに前記スペクトル傾きの度合いによって決定されるスペクトル減算係数を乗じる処理と、前記入力ス

ペクトルから前記スペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクションスペクトルを得る処理と、前記サブトラクションスペクトルのクリッピングを行って音声スペクトルを求める処理と、前記音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑化して修正することにより、前記雑音成分が抑圧された出力音声信号を得る処理とをコンピュータに実行させるためのプログラムまたは該プログラムを格納したコンピュータ読み取り可能な記録媒体。

【 0 0 6 1 】

【発明の効果】

以上説明したように、本発明によれば入力スペクトルから推定雑音スペクトルを減算した後のスペクトルをクリッピングした後に周波数軸上や時間軸上で平滑化して修正することにより、雑音成分を抑圧しつつ、出力音声信号のスペクトルを本来の音声スペクトルの概形に近づけることが可能となる。また、スペクトル減算係数を推定雑音スペクトルの形状に基づいて算出することによって、より正確なスペクトル減算を行い、良好な雑音抑圧効果が得られる。さらに、本発明の雑音抑圧処理を音声認識処理の前処理として用いることにより、雑音環境下において高い認識率を達成できる。

【図面の簡単な説明】

【図 1】 スペクトルサブトラクション法の第 1 の問題点を説明するための入力スペクトルと推定雑音スペクトル及び出力スペクトルの例を示す図

【図 2】 クリーン条件でのスペクトルサブトラクション法による出力スペクトルを示す図

【図 3】 スペクトルサブトラクション法の第 2 の問題点を説明するための入力スペクトルと推定雑音スペクトル及び出力スペクトルの例を示す図

【図 4】 スペクトルサブトラクション法の第 3 の問題点を説明するための高域振幅が大きい本来の雑音スペクトルと推定雑音スペクトルについて示す図

【図 5】 スペクトルサブトラクション法の第 3 の問題点を説明するための高域振幅が小さい本来の雑音スペクトルと推定雑音スペクトルについて示す図

【図 6】 本発明の第 1 の実施形態に係る雑音抑圧装置の構成を示すブロック



図

【図 7】第 1 の実施形態における雑音抑圧処理の流れを示すフローチャート

【図 8】第 1 の実施形態において音声スペクトルを周波数軸上で近傍のスペクトルを用いて修正した場合の修正前と修正後のスペクトル及びクリーン条件のスペクトルを示す図

【図 9】第 1 の実施形態において音声スペクトルを所定の関数と畳み込みを行うことで修正した場合の修正前と修正後のスペクトル及びクリーン条件のスペクトルを示す図

【図 1 0】第 1 の実施形態において音声スペクトルを時間軸上で近傍のスペクトルを用いて修正した場合の修正前と修正後のスペクトルを示す図

【図 1 1】本発明の第 2 の実施形態に係る雑音抑圧装置の構成を示すブロック図

【図 1 2】第 2 の実施形態における雑音抑圧処理の流れを示すフローチャート

【図 1 3】本発明の第 3 の実施形態に係る雑音抑圧装置の構成を示すブロック図

【図 1 4】第 3 の実施形態における雑音抑圧処理の流れを示すフローチャート

【図 1 5】本発明の第 4 の実施形態に係る音声認識装置の構成を示すブロック図

【符号の説明】

- 1 1 … 音声入力端子
- 1 2 … 周波数分析部
- 1 3 … 雑音スペクトル推定部
- 1 4 … スペクトル減算係数格納部
- 1 5 … 乗算器
- 1 6 … 減算器
- 1 7 … クリッピング部
- 1 8 … スペクトル修正部

1 9 … 音声出力端子

2 1 … スペクトル傾き算出部

2 2 … スペクトル減算係数算出部

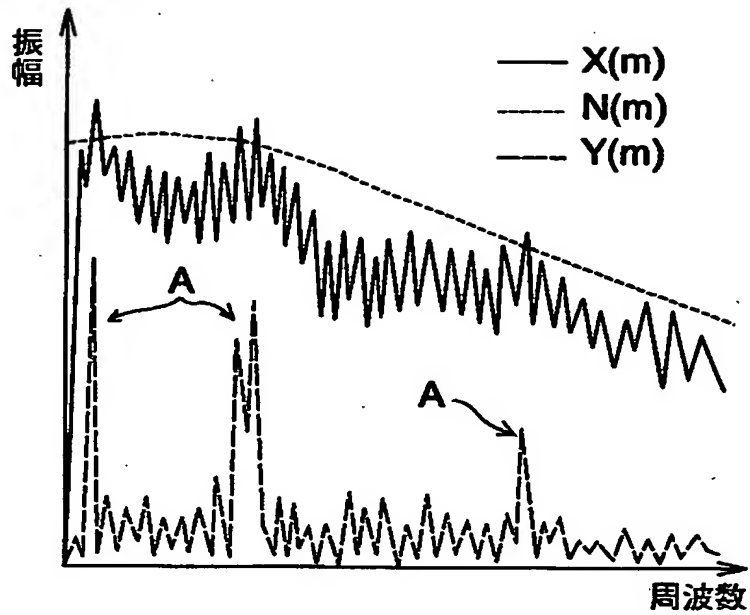
3 1 … 雑音抑圧部

3 2 … 音声認識部

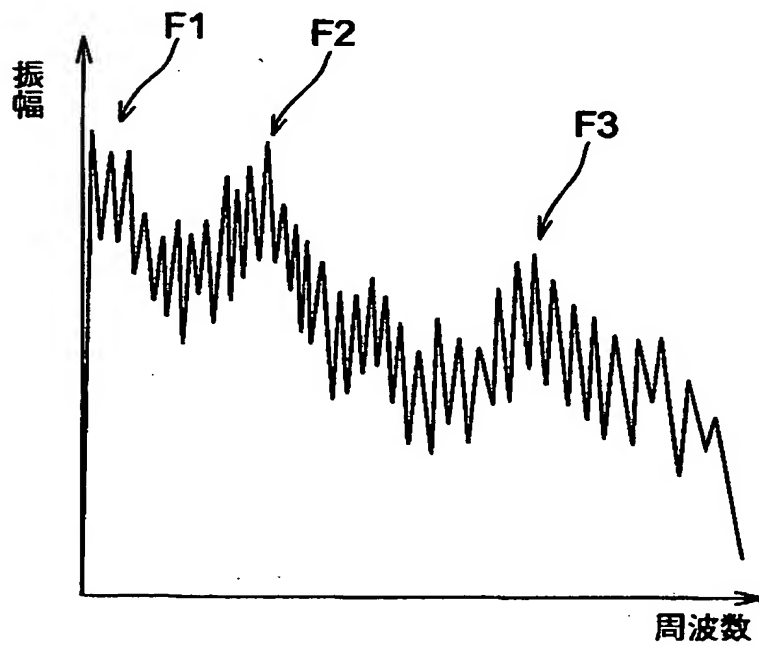
【書類名】

図面

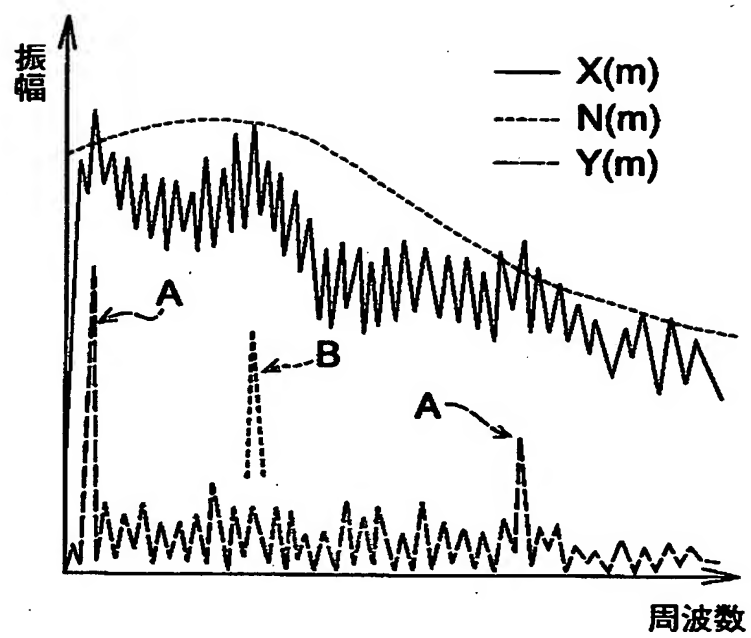
【図 1】



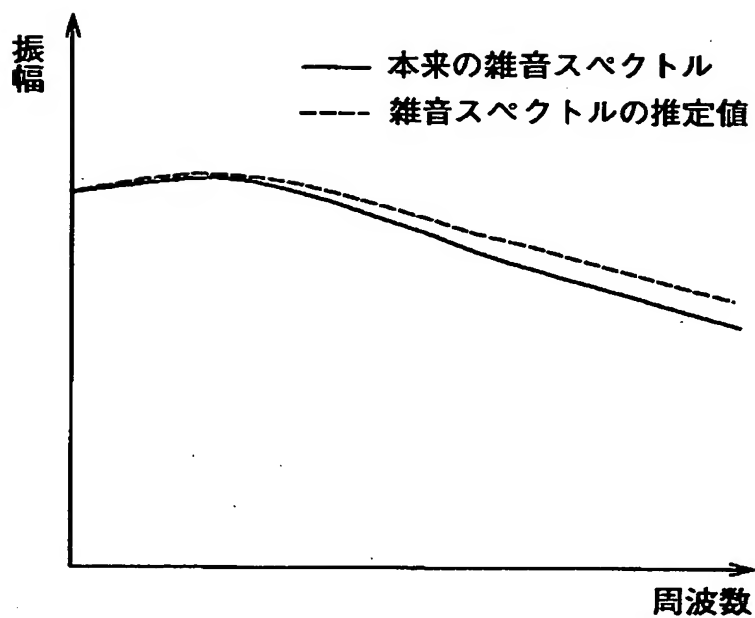
【図 2】



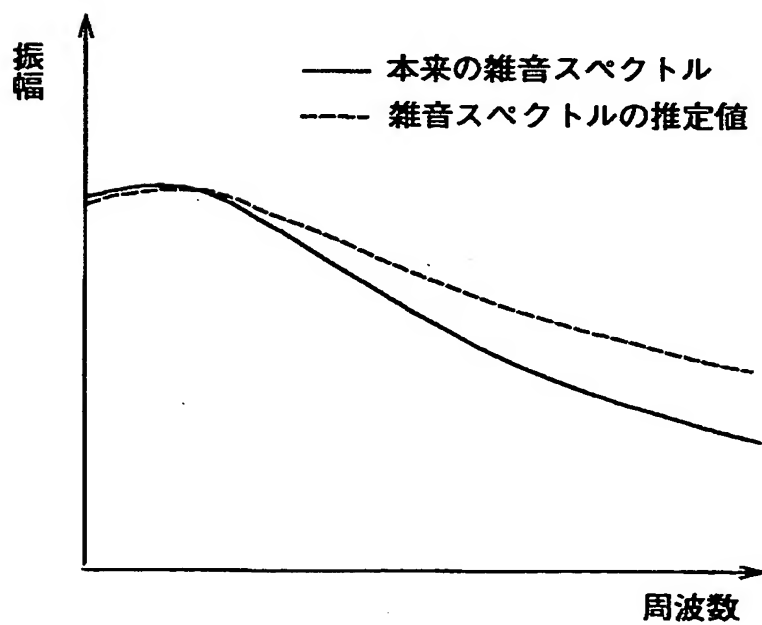
【図3】



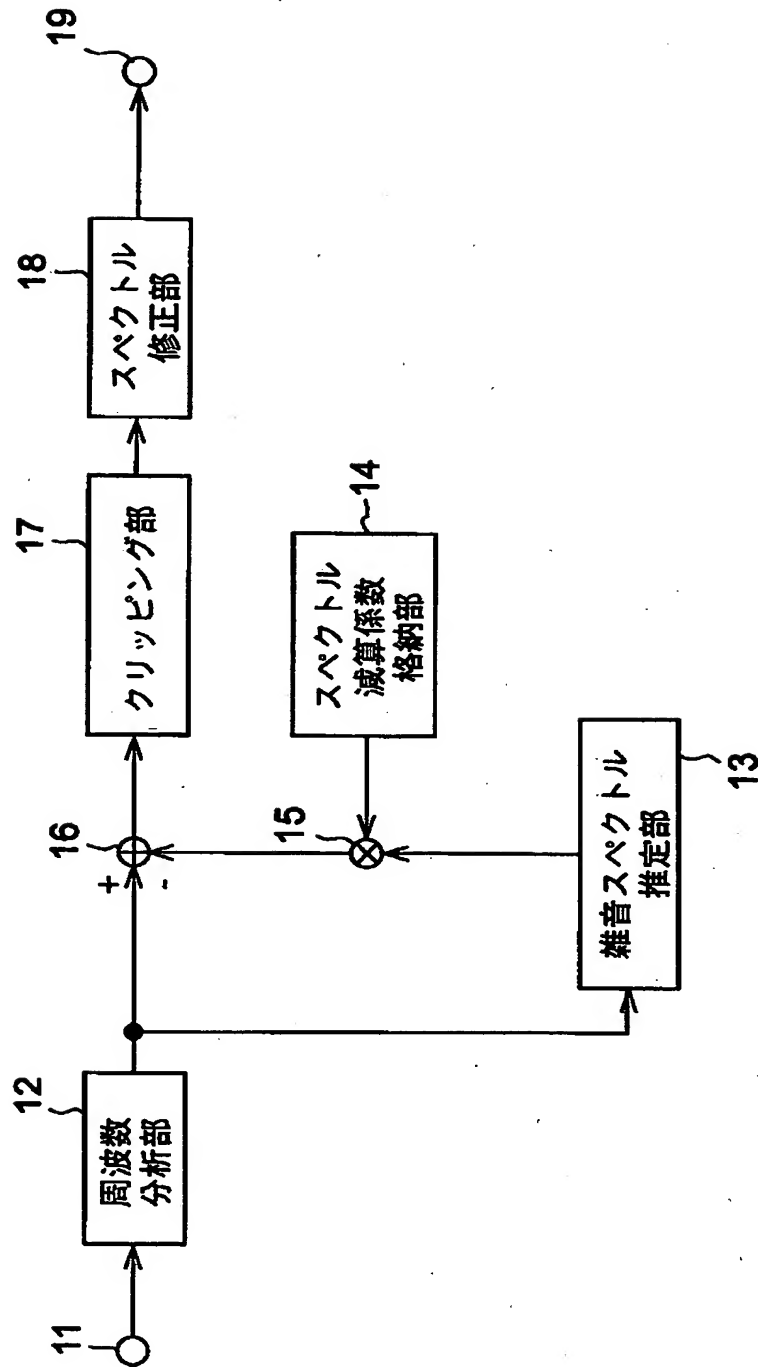
【図 4】



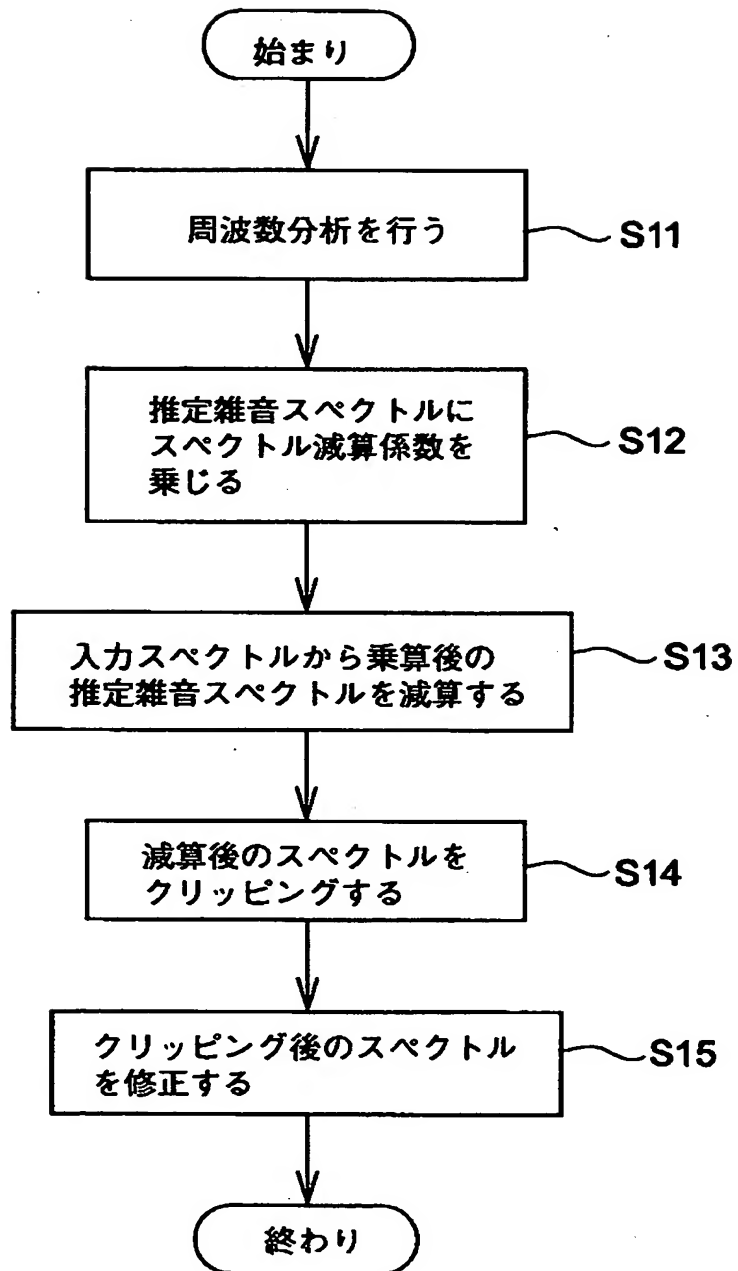
【図 5】



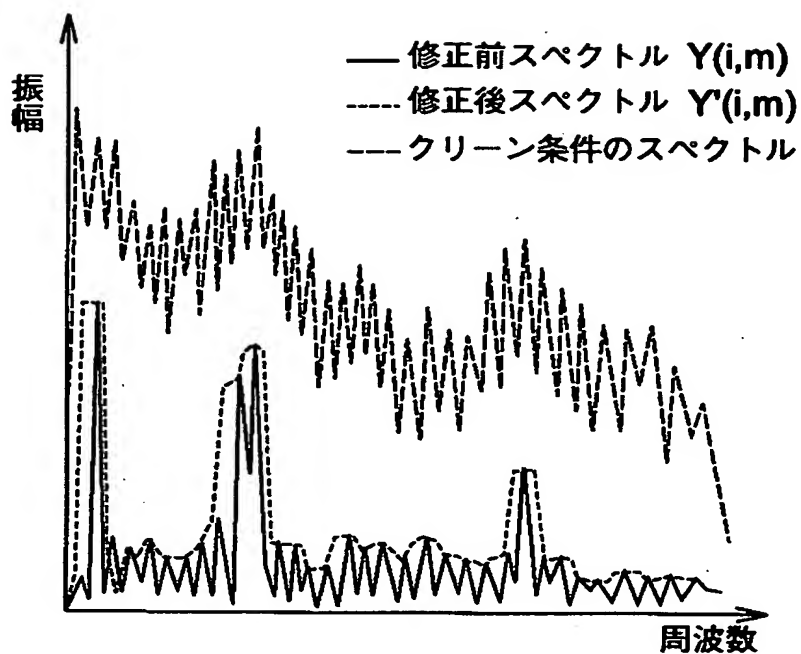
【図 6】



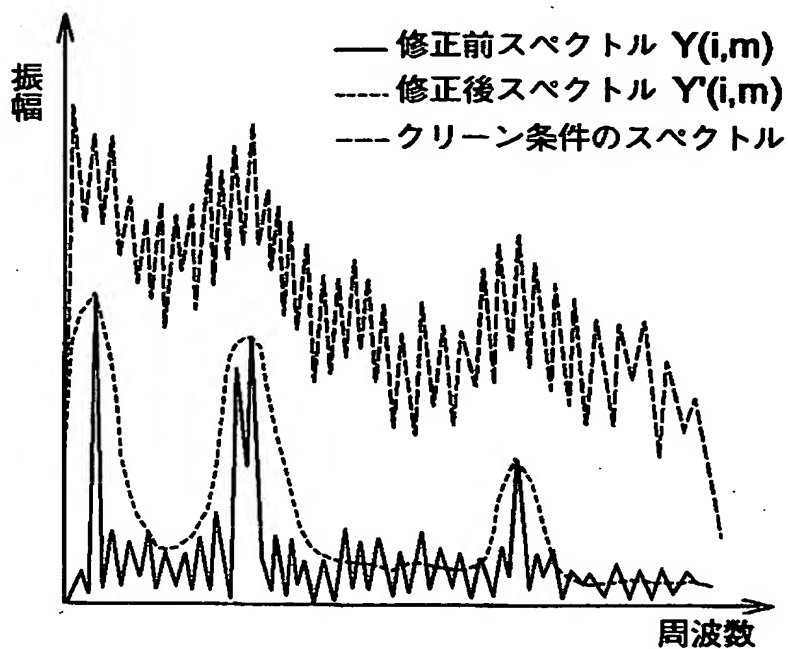
【図7】



【図8】

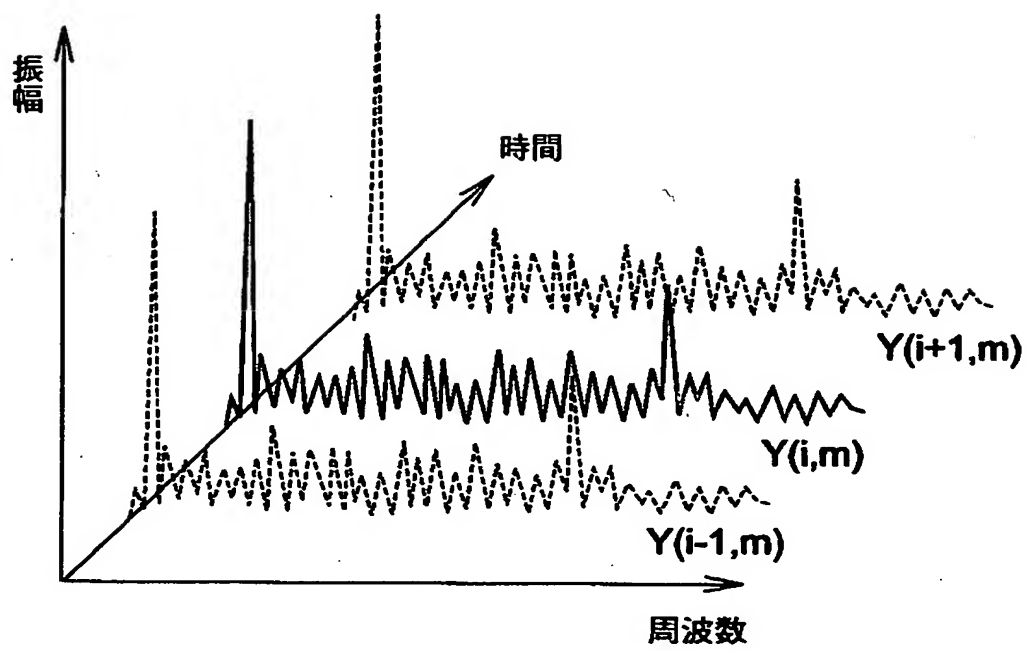


【図9】

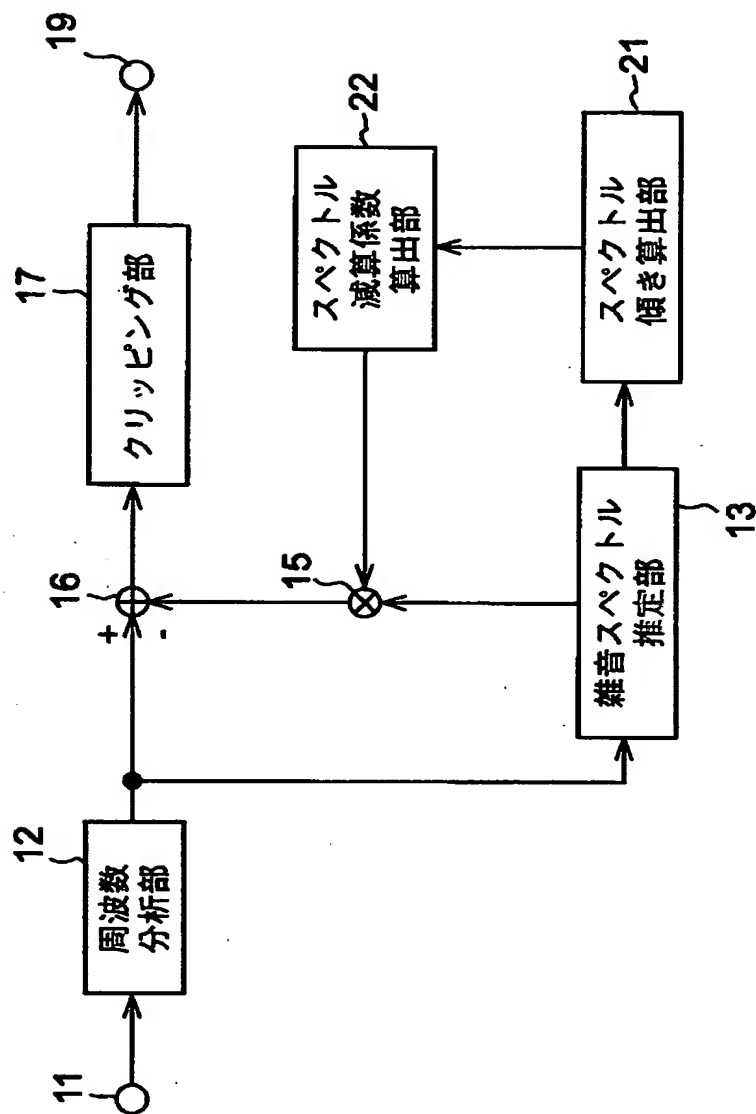




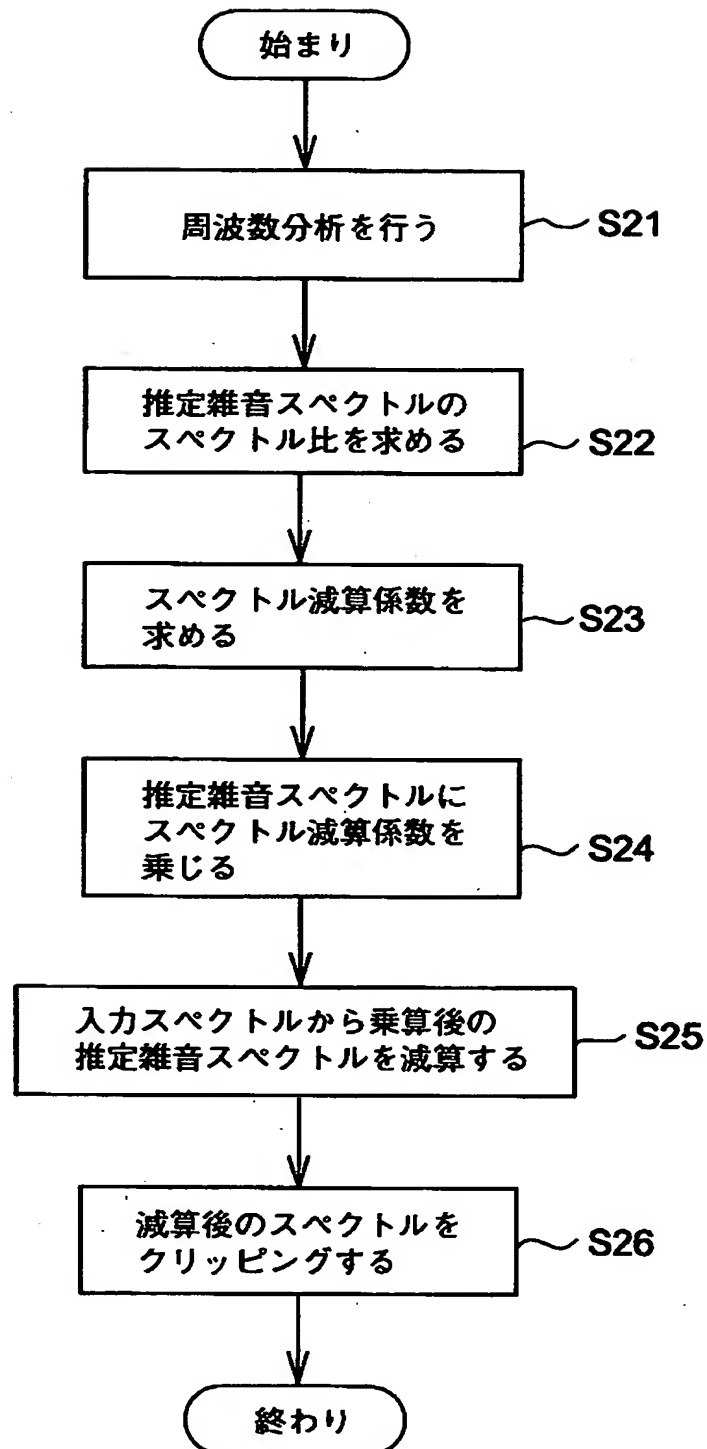
【図10】



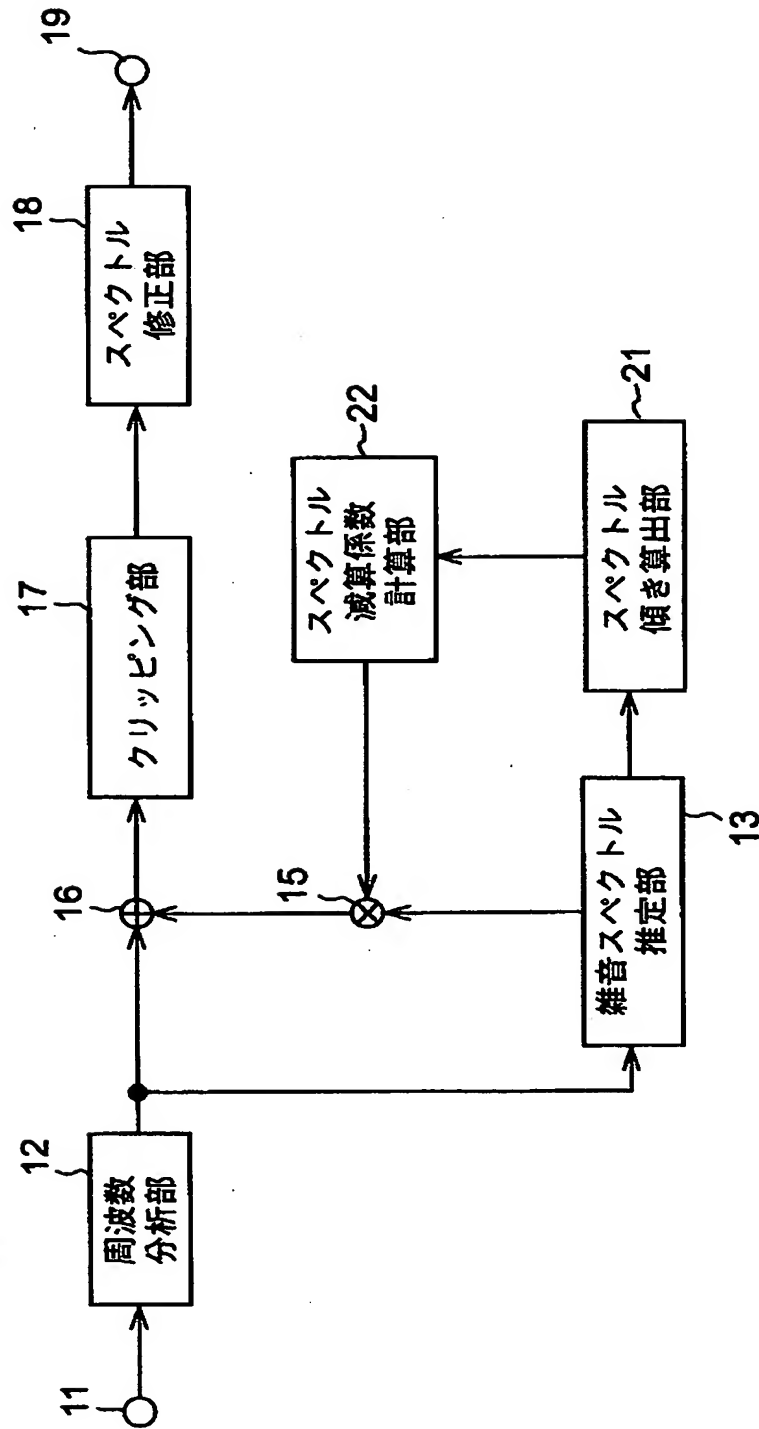
【図11】



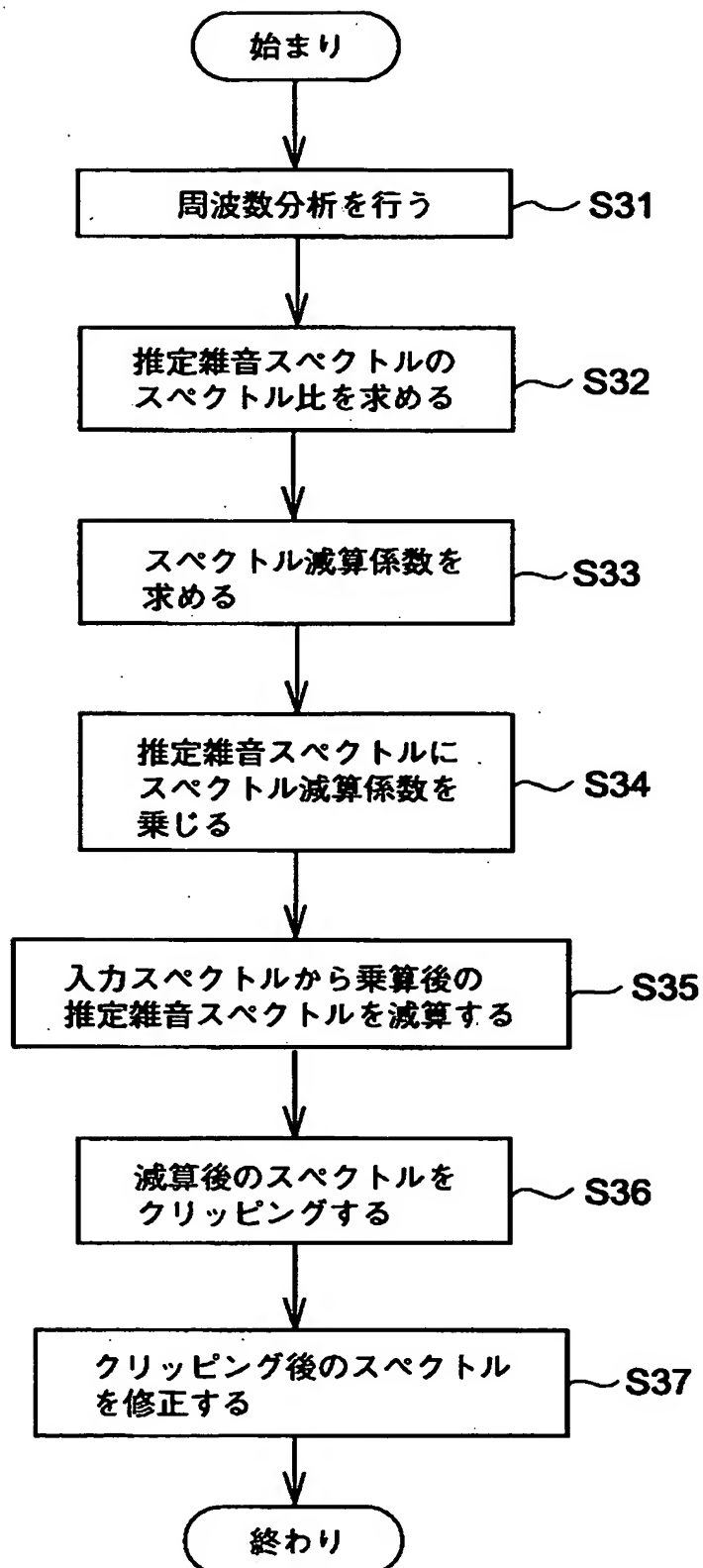
【図12】



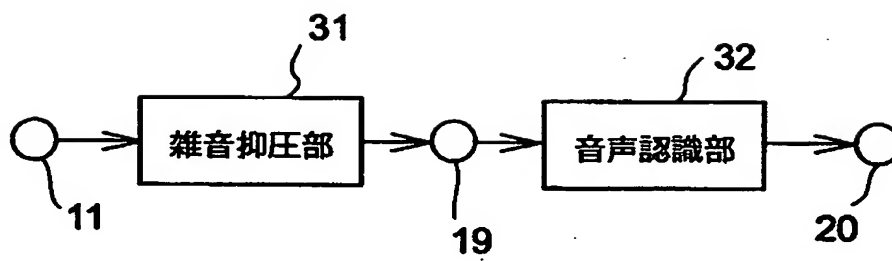
【図 13】



【図14】



【図 1 5】



【書類名】                      要約書

【要約】

【課題】 入力音声信号に含まれる雑音成分を音声信号のスペクトルを損なうことなく抑圧できる雑音抑圧装置を提供する。

【解決手段】 入力端子 1 1 への入力音声信号を所定フレーム長で周波数分析して入力スペクトルを得る周波数分析部 1 2 と、雑音成分のスペクトルを推定する雑音スペクトル推定部 1 3 と、推定雑音スペクトルにスペクトル減算係数を乗じる乗算器 1 5 と、入力スペクトルからスペクトル減算係数を乗算した後の推定雑音スペクトルを減じてサブトラクションスペクトルを得る減算器 1 6 と、サブトラクションスペクトルのクリッピングを行って音声スペクトルを求めるクリッピング部 1 7 と、音声スペクトルを周波数軸上及び時間軸上の少なくとも一方について平滑化して修正することにより雑音成分が抑圧された出力音声信号を得るスペクトル修正部 1 8 とを有する。

【選択図】                      図 6

出 願 人 履 歴 情 報

識別番号 [000003078]

1. 変更年月日 1990年 8月22日  
[変更理由] 新規登録  
住 所 神奈川県川崎市幸区堀川町72番地  
氏 名 株式会社東芝
2. 変更年月日 2001年 7月 2日  
[変更理由] 住所変更  
住 所 東京都港区芝浦一丁目1番1号  
氏 名 株式会社東芝